

ARTICLE

Knowledge graph aids comprehensive explanation of drug and chemical toxicity

Yun Hao¹ | Joseph D. Romano^{2,3} | Jason H. Moore⁴

¹Genomics and Computational Biology (GCB) Graduate Program, University of Pennsylvania, Philadelphia, Pennsylvania, USA

²Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, Pennsylvania, USA

³Center of Excellence in Environmental Toxicology, University of Pennsylvania, Philadelphia, Pennsylvania, USA

⁴Department of Computational Biomedicine, Cedars-Sinai Medical Center, Los Angeles, California, USA

Correspondence

Jason H. Moore, Department of Computational Biomedicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA.

Email: jason.moore@csmc.edu

Abstract

In computational toxicology, prediction of complex endpoints has always been challenging, as they often involve multiple distinct mechanisms. State-of-the-art models are either limited by low accuracy, or lack of interpretability due to their black-box nature. Here, we introduce AIDTox, an interpretable deep learning model which incorporates curated knowledge of chemical-gene connections, gene-pathway annotations, and pathway hierarchy. AIDTox accurately predicts cytotoxicity outcomes in HepG2 and HEK293 cells. It also provides comprehensive explanations of cytotoxicity covering multiple aspects of drug activity, including target interaction, metabolism, and elimination. In summary, AIDTox provides a computational framework for unveiling cellular mechanisms for complex toxicity endpoints.

Study Highlights**WHAT IS THE CURRENT KNOWLEDGE ON THIS TOPIC?**

In computational toxicology, prediction of complex endpoints has always been challenging, as they often involve multiple distinct mechanisms. State-of-the-art models are either limited by low accuracy, or lack of interpretability due to their black-box nature.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

This study addressed the black-box nature of machine learning applications in the field. In doing so, it significantly improved the interpretability of toxicity prediction.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

This study introduced AIDTox, an interpretable deep learning model which incorporates curated knowledge of chemical-gene connections, gene-pathway annotations, and pathway hierarchy. AIDTox can provide comprehensive explanations of complex toxicity endpoints covering multiple aspects of drug activity, including target interaction, metabolism, and elimination. In doing so, it provides a computational framework for unveiling mechanisms of toxicity.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *CPT: Pharmacometrics & Systems Pharmacology* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?

AIDTox can be applied to drug discovery research in two perspectives. First, AIDTox can be applied to prioritize compounds for experimental safety testing, from large-scale chemical libraries that contain thousands of compounds. Second, AIDTox can be applied to prioritize target proteins and pathways responsible for toxicity endpoints, in order to generate new hypothesis for mechanistic investigation.

INTRODUCTION

In silico toxicity assessment is becoming a critical step in drug development owing to sustained advances of machine learning techniques.¹ A prevalent approach for toxicity prediction is quantitative structure–activity relationship (QSAR) modeling, which links a toxicity outcome to structural properties of compounds.² Coupled with millions of data points generated by large-scale toxicity testing,^{3,4} various QSAR models have been proposed to predict in vitro and in vivo endpoints.^{5–13} Whereas some have achieved decent predictive performance, almost none can overcome the trade-off between accuracy and interpretability.¹⁴ Under state-of-the-art models, it remains challenging to explain the toxicity outcomes of a compound with cellular activities involving target proteins, specific pathways, and biological processes. This limitation has raised substantial concerns among experimental toxicologists, calling for prediction models that can provide insight into cellular mechanisms of toxicity.

Recent developments in visible neural networks (VNNs) provide a solution to the issue.^{15–17} In contrast to black-box neural networks, connections within VNNs are guided by curated knowledge from pathway ontologies. Specifically, the fully connected structure is replaced by an interpretable ontological hierarchy that connects input gene features to output response via hidden pathway modules. In a previous study, we developed a VNN model—named DTox—for predicting compound response to toxicity assays.¹⁸ Importantly, DTox advances an innovative interpretation framework that identifies network paths connecting genes and pathways for explaining the toxicity outcome of compounds. We demonstrated that DTox can achieve the same level of predictive performance as state-of-the-art models with a significant improvement in interpretability, as the identified paths can be linked to well-established mechanisms of toxicity. However, one major limitation of DTox is that the input feature profile is derived from structure-based binding prediction models. Although these models ensure the wide applicability of DTox, they are vulnerable to prediction errors and data scarcity, causing exclusion of certain

genes from the feature space. To address this limitation, we used curated knowledge from the toxicology-focused graph knowledge base ComptoxAI¹⁹ to refine the input feature profile of DTox. ComptoxAI provides extensive profiling of chemical-gene connections across multiple gene categories, which can be used for feature profile construction. Hence, the resulting model—named AIDTox (ComptoxAI + DTox)—contains many novel gene features with active roles in cellular mechanisms of toxicity, including tubulin proteins that regulate apoptosis signaling, cytochrome P450 enzymes that are mainly responsible for drug metabolism, and transporters that participate in drug elimination. We believe AIDTox will facilitate the generation of new hypotheses for mechanistic investigation. Our code can be accessed openly at <https://github.com/yhao-compbio/AIDTox>. We also provide a glossary of machine learning terms mentioned in this paper (Supplementary Text).

METHODS

Processing cell viability screening datasets for model training

The Tox21 datasets³ contain screening results indicating the response of in vitro toxicity assays to compounds of interest. We focused our analyses on two cytotoxicity assays: HEK293 and HepG2, for each of which at least 500 compounds are available for model training. We first removed compounds with inconclusive or ambiguous screening results, then grouped the remaining compounds into active (cytotoxicity detected in at least one experimental concentration) and inactive (no cytotoxicity detected in any concentration; Table S1).

Extracting chemical-gene connections from ComptoxAI for model construction

ComptoxAI^{19,20} is a comprehensive graph knowledge base that contains curated relationships among chemicals,

genes, assays, and many other entities. It contains two types of relationships linking compound and gene nodes: physical binding (with protein product) and expression-alteration (upregulation/downregulation). We extracted both types for chemicals present in the Tox21 datasets. We also constructed a hybrid type by combining the two types of relationships. We used the extracted relationships to assemble binary compound-gene matrices as input feature profiles.

To perform dimensionality reduction, we implemented a ReliefF-based method—namely, MultiSURF²¹—for ranking genes by relevance to the assay outcome of interest. MultiSURF takes in a labeled feature dataset (assay outcome as the label), and ranks all features based on differences among neighboring instances. Specifically, intraclass differences will contribute negatively to feature relevance, whereas interclass differences will contribute positively. A benchmark study showed that MultiSURF outperformed other methods in detecting genotype–phenotype associations.²¹ We selected the top 100, 200, 300, 400, and 500 ranked genes of each dataset to proceed with the following analysis. Model selection by training loss was carried out to identify the optimal gene feature space (Table S2).

Constructing VNN with selected gene features and Reactome pathway hierarchy

DTox is a VNN model embedded with the Reactome pathway hierarchy that comprises root biological processes, child–parent pathway relations, and gene–pathway annotations (downloaded in August 2019).²² Each pathway is embedded as a module consisting of hidden neurons. AIDTox inherits the basic structure of DTox while modifying input layers to consist of the gene features selected by MultiSURF. Guided by gene–pathway annotations, the gene features are connected to modules representing lowest-level pathways. These modules are then connected to deeper layers based on child–parent pathway relations until root biological processes are reached. Finally, the root biological processes are connected to an output layer representing the assay outcome. To trim the network's scale and prevent overfitting, we made the root biological process a hyperparameter of AIDTox. We selected four processes (“gene expression,” “immune system,” “metabolism,” and “signal transduction”) and all possible combinations among them for the tuning process (15 values in total). These four processes were selected due to their broad coverage and direct involvement in cellular mechanisms of toxicity.

AIDTox also inherits the hybrid loss function of DTox combining both root and auxiliary loss terms:

$$\text{BCELoss}(y_{\text{root}}, y) + \alpha \sum_p \beta_p \text{BCELoss}(y'_{p-\text{auxi}}, y) + \lambda \|W\|_2.$$

The first term, namely root loss, denotes the binary cross entropy of final output y_{root} . The second term, namely auxiliary loss, denotes the binary cross entropy of the auxiliary scalar $y'_{p-\text{auxi}}$ from each pathway module, with factor α ($=0.5$) balancing the root and auxiliary terms, and factor β_p (computed as the inverse of pathway count within the corresponding hidden layer) normalizing terms across hidden layers. The third term denotes L_2 regularization with coefficient λ ($=1e^{-4}$). The incorporation of auxiliary loss terms can prevent gradients from vanishing in the lower hierarchy and facilitates learning new patterns from individual pathways.

Learning optimal AIDTox model for cytotoxicity prediction and interpretation

AIDTox adopts the same training scheme as DTox, including a training/testing/validation split by ratio of 7:1:2, optimization using the Adam algorithm²³ with mini-batch size of 32, a learning rate of 0.001, early stopping criterion with “patience” of 20, and hyperparameter tuning by grid search. We evaluated the performance of the optimal AIDTox model on the held-out validation set (Table S3), and compared it with three existing models: (i) the optimal DTox model, (ii) an optimal QSAR model based on a random forest classifier (derived from tuning 6 hyperparameters with 2800 combinations in total, listed in Table S4), and (iii) an optimal QSAR model based on a gradient boosting classifier (derived from tuning five hyperparameters with 3000 combinations in total, listed in Table S4). For the two QSAR models (ii and iii), we adopted 166-bit binary MACCS fingerprints to quantify structural properties of compounds, which covers most of the interesting physicochemical features for drug discovery.²⁴ We also implemented DTox's model interpretation framework to identify paths from the VNN that can explain compound cytotoxicity (Table S5).

RESULTS

AIDTox uses curated chemical-gene connections to construct VNN

AIDTox predicts and explains compound response to toxicity assays with a knowledge-guided VNN. The knowledge incorporated in AIDTox comprises chemical-gene connections from ComptoxAI, as well as gene–pathway annotations and child–parent pathway relationships from Reactome (Figure 1a). Connections within the VNN are constrained to gene–pathway connections (input to first

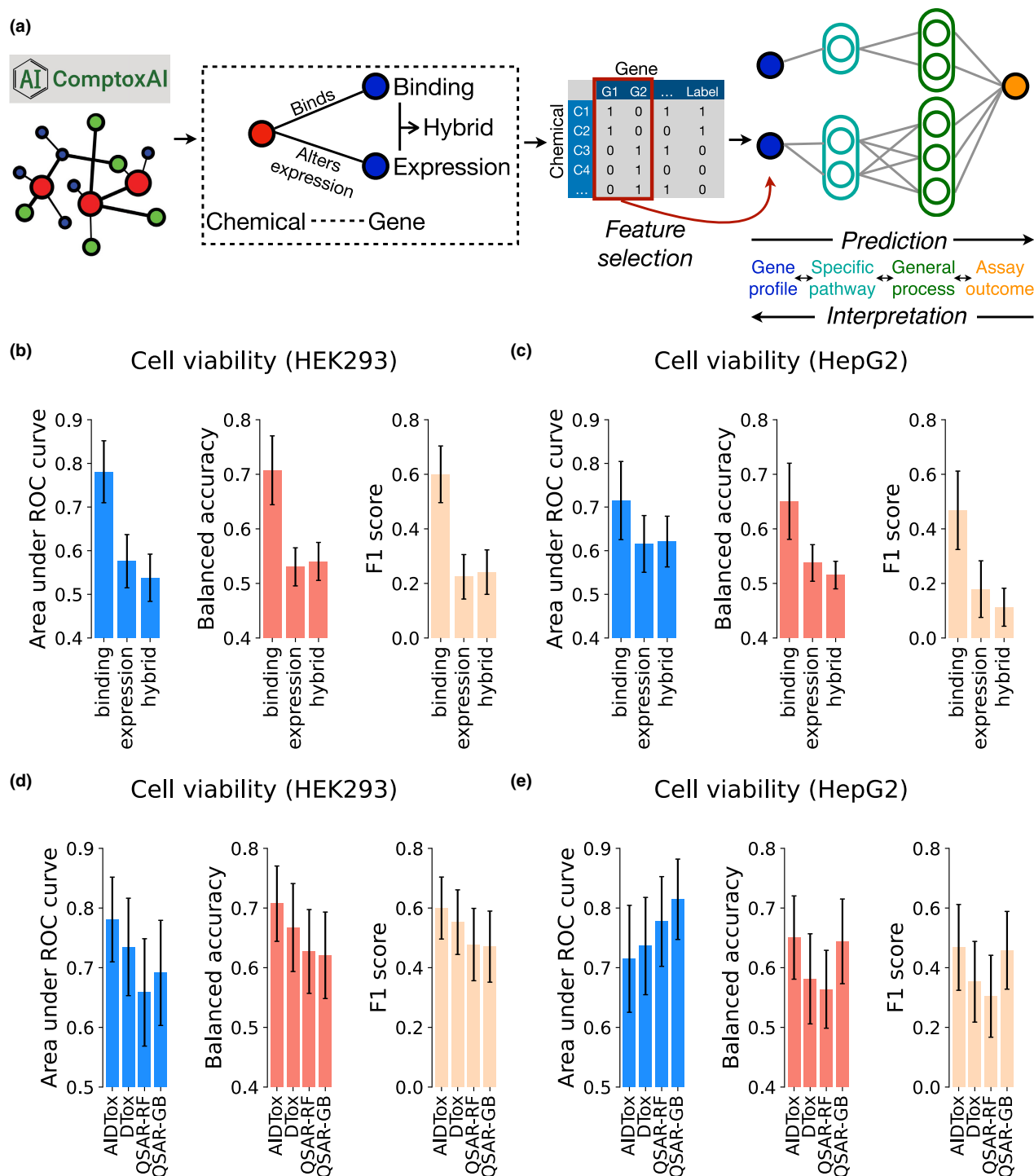


FIGURE 1 Incorporating curated chemical-gene connections into VNN for toxicity prediction with AIDTox. (a) Three types of chemical-gene connections (binding, expression, and hybrid) are extracted from ComptoxAI to construct the input feature profile of AIDTox. Feature selection is implemented to identify the top gene features predictive of the outcome of interest. The selected profile is fed into a VNN, whose structure is guided by Reactome pathway hierarchy. Specific pathways and general processes are coded as modules by hidden neurons. Barplots showing the comparison of validation performance across three connection types in two cell viability datasets: HEK293 (b) and HepG2 (c). Performance is measured by three metrics: area under ROC curve, balanced accuracy, and F1 score, with error bar showing the 95% confidence interval. Barplots showing the comparison of validation performance across four models in two cell viability datasets: HEK293 (d) and HepG2 (e). Three other models are considered: (i) our previous DTTox model with inferred target profile as input, (ii) QSAR model by random forest with chemical fingerprint as input, and (iii) QSAR model by gradient boosting with chemical fingerprint as input. For area under ROC curve and balanced accuracy scores, the “no predicative value” level is 0.5. QSAR, quantitative structure–activity relationship; ROC, receiver operating characteristic; VNN, visible neural network.

hidden layer) and child–parent pathway relations (after first hidden layer). In this study, we focused on two cell viability assays measuring compound cytotoxicity in HEK293 and HepG2 cells. Both datasets contain 1367 compounds with connections in ComptoxAI, including 432 cytotoxic and 935 non-cytotoxic compounds in HEK293 cells, and 293 cytotoxic and 1074 non-cytotoxic compounds in HepG2 cells. Among these compounds, 617 (45%) are connected to 991 genes with physical binding evidence, 1237 (90%) are connected to 8723 genes with expression-alteration evidence, and 1367 (100%) are connected to 8735 genes with both types of evidence. In all three cases, the number of gene features is much greater than the number of compound samples. To prevent overfitting, we reduced feature dimensions by selecting the top 100–500 predictive genes during model construction. In general, as more predictive genes were incorporated, we observed an improvement in the resulting model performance (decline in training loss) until 400 (Figure S1). Therefore, we proceeded with models built using the top 400 predictive genes.

Chemical-gene binding connections result in the best performing models

Using held-out validation sets of HEK293 and HepG2 cell viability data, we first compared the performance of models derived from three distinct connection types. In both datasets, models derived from binding connections significantly outperform the other types (Figure 1b,c; Table S2), as we observed no overlaps between the 95% confidence intervals of compared metrics (area under the receiver operating characteristic curve [AUROC], balanced accuracy, and F1-score) except for AUROC on the HepG2 viability dataset. We then compared the performance of AIDTox models derived from binding connections against three other classification algorithms, including our previous DTTox model (derived from predicted compound–target interactions) and two QSAR models (derived from

quantified structural properties). In both datasets, the 95% confidence intervals of performance metrics are highly overlapped among compared algorithms (Figure 1d,e; Table S3). Therefore, AIDTox achieved the same level of predictive performance as these well-established classification algorithms.

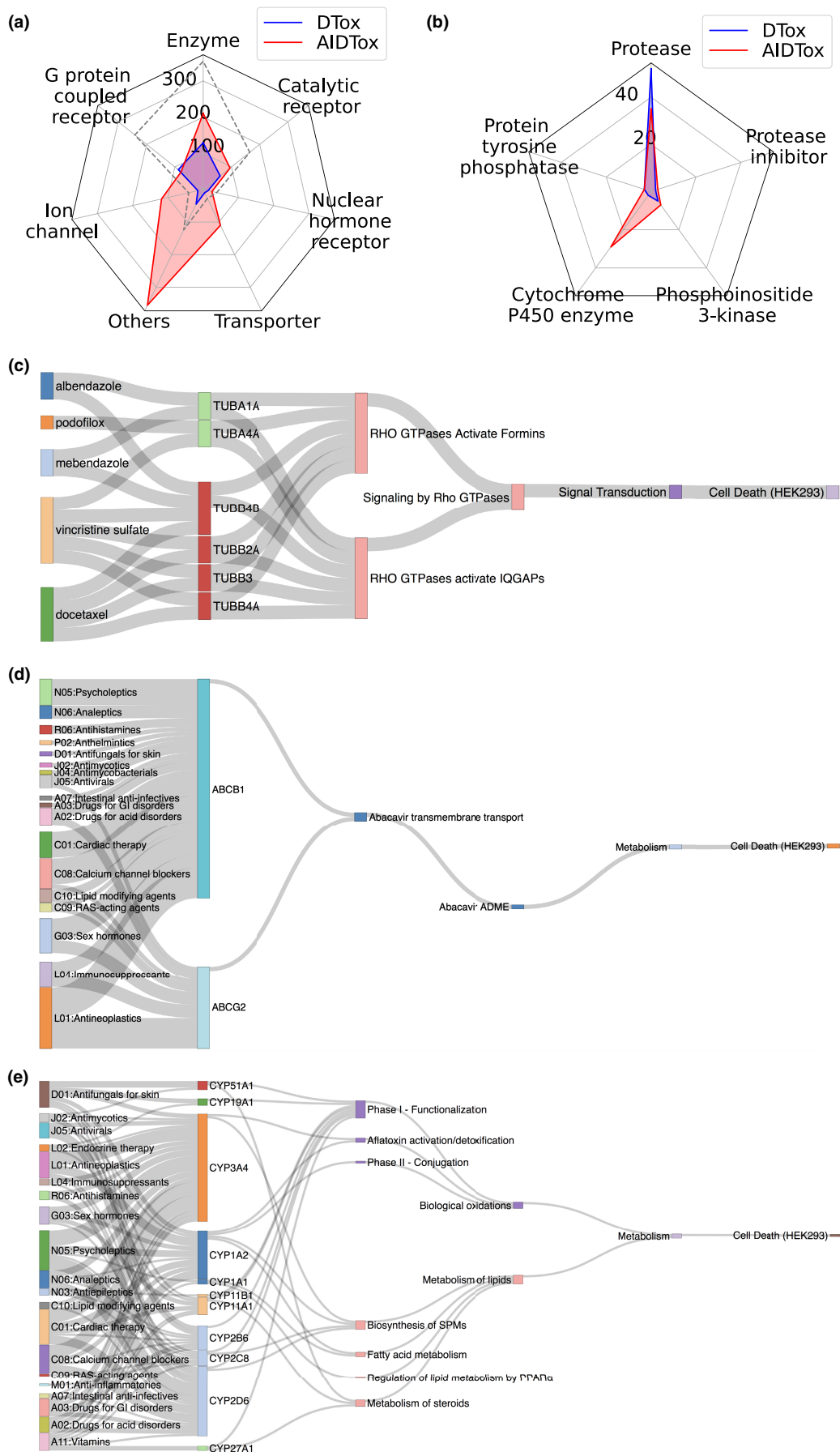
AIDTox models benefit from comprehensive gene feature space

A fundamental advantage of AIDTox over DTTox comes from its enlarged gene feature space, providing an extensive profiling of the cellular activities of compounds. Overall, there is an increase from 361 genes in DTTox to 991 genes in AIDTox, a 2.7-fold increase. Comparing the composition of gene features, we discovered that the increase in AIDTox is mainly driven by three target categories: ion channels, transporters, and others (Figure 2a). One example of the “others” category is the tubulin protein superfamily (Figure 2c), which polymerize into microtubules, a primary component of cytoskeleton involved in cell division. Due to the lack of binding data, tubulin proteins were not present in the DTTox model. In ComptoxAI, the two families of tubulin proteins (α and β) are connected to antiparasitics (albendazole and mebendazole) and antineoplastics (podofilox, vincristine sulfate, and docetaxel). Our AIDTox model further connects these drugs to HEK293 cytotoxicity via Rho GTPase signaling by effectors, such as formins and IQGAPs, which have been shown to regulate apoptosis in HEK293 cells.²⁵

New features in AIDTox are essential in drug metabolism and elimination processes

We observed a disproportionate 11-fold increase (from 10 to 110) among transporters in the gene feature space of AIDTox. For instance, AIDTox connects drugs within nine

FIGURE 2 Comprehensive explanation of HEK293 cytotoxicity with new features in AIDTox. (a) Radar plot showing the comparison of gene category distributions among the features of DTTox (blue solid line) and AIDTox (red solid line). The gray dashed line shows the hypothetical of a proportionate increase from DTTox to AIDTox. (b) Radar plot showing the comparison of enzyme subcategory distributions among the features of DTTox (blue solid line) and AIDTox (red solid line). (c) Sankey diagram showing the AIDTox explanation of HEK293 cytotoxicity for drugs targeting tubulin proteins. The paths (connecting drugs to HEK293 cell death) shown in the diagram are identified from the full network of VNN model by the AIDTox interpretation framework. Connections in the VNN are informed by ComptoxAI (chemical–gene) and Reactome (gene–pathway and child–parent pathway). Tubulin proteins are grouped and colored by the family. (d) Sankey diagram showing the AIDTox explanation of HEK293 cytotoxicity via ATP-binding cassette transporters (similar to c). Drugs are grouped and colored by the ATC subclass (first three digits of the ATC code). (e) Sankey diagram showing the AIDTox explanation of HEK293 cytotoxicity via cytochrome P450 enzymes (similar to c). Drugs are grouped and colored by the ATC subclass (first three digits of the ATC code). Cytochrome P450 enzymes are grouped and colored by the family. Metabolic pathways are grouped and colored by the general metabolic process they belong to (“Biological oxidations” or “Metabolism of lipids”). ADME, absorption, distribution, metabolism, and excretion; ATC, Anatomic Therapeutic Chemical.



ATC classes (18 subclasses) to HEK293 cytotoxicity via two new features: transporters *ABCB1* and *ABCG2* (Figure 2d). These two ATP-binding cassette transporters are mainly responsible for pumping drugs out of the cell; thus they play a critical role in drug elimination. Therefore, inhibition of transporter-mediated elimination by drugs may prolong their cytotoxic effect on HEK293 cells.²⁶

Although we did not observe a disproportionate increase among the general enzyme category, we did observe a 14.5-fold increase (from 2 to 29) among one subcategory: the cytochrome P450 enzymes (CYPs; Figure 2b). For instance, AIDTox connects drugs within nine ATC classes (20 subclasses) to HEK293 cytotoxicity via 20 CYPs of seven families (Figure 2e). These CYPs are the major enzymes involved in drug metabolism as they participate in metabolic pathways, such as “biological oxidation” and “metabolism of lipids.” Meanwhile, CYP-mediated metabolic processes are the main source of reactive oxygen species, which can lead to cellular oxidative stress and trigger apoptosis/necrosis.²⁷

DISCUSSION

Rich knowledge in ComptoxAI provides accurate and extensive profiling of chemical-gene connections. Here, we explored the incorporation of these connections into knowledge-guided deep learning models for predicting and explaining compound cytotoxicity. We considered three types of connections for the task: physical binding, expression-alteration, and a hybrid type. Models derived from binding connections exhibit the best predictive performance, because physical binding is stronger evidence of direct interaction compared to expression alteration. In contrast to our previous work DTox, the new AIDTox model uses curated knowledge for generating input feature profile, thus is not prone to errors from binding prediction models. AIDTox is not restrained by the availability of compound-target binding data. The feature space of AIDTox comprises many genes with insufficient binding data, including ion channels, transporters, and CYP enzymes. These categories exhibit central roles in drug metabolism and elimination processes. Accordingly, they become an asset for prediction and explanation of complex toxicity outcomes, which may be triggered by multiple cellular mechanisms. For instance, AIDTox was able to recapitulate some target proteins and pathways responsible for cell death, such as tubulins that regulate apoptosis via IQGAP-mediated Rho GTPase signaling,²⁵ as well as CYPs that regulate apoptosis via oxidative stress.²⁷ In another example, AIDTox was able to connect dasatinib, a leukemia drug, to HEK293 cytotoxicity via multiple

aspects of drug activity, including MAPK14-mediated regulation of apoptosis, CYP1A2/CYP3A4-mediated drug metabolism, and *ABCB1*/*ABCG2*-mediated drug elimination (Figure S2). It is also worth noticing the high interpretability of AIDTox is achieved without loss of accuracy. Therefore, we anticipate AIDTox to be applied in both prioritizing compounds for safety testing and generating new hypothesis for mechanistic investigation.

Despite these highlights, AIDTox did not significantly outperform existing methods in cytotoxicity prediction. This can be attributed to the sharp decline in sample size, as the profiling of chemical-gene connections by ComptoxAI only applies to well-studied compounds. A smaller sample size in turn leads to wider confidence intervals of model performance metrics. Consequently, even though we observed a moderate increase in the performance metrics, the increase is well within the 95% confidence interval. In the future, we expect the performance of AIDTox to be enhanced after chemical-gene connections become available for more compounds. We also acknowledge the recent development of graph neural network-based link prediction algorithms, which can fill in missing connections for under-studied compounds. Such algorithms may help us increase the sample size for AIDTox training and enhance model performance. In addition to increasing sample size, we think the incorporation of context-specific knowledge, such as cell line-specific chemical-gene connections, may further enhance the performance of AIDTox. Although AIDTox achieved some success in predicting cytotoxicity, further work is required to refine the model for risk prediction in vivo, given that cytotoxicity only represents an extreme endpoint in vitro. Many factors need to be taken into account regarding the model refinement, including the difference in target protein expression between immortal cell lines and primary cell types, as well as the limited training data from in vivo screening. Whereas it may be challenging to directly fit a knowledge-guided deep learning model with the limited in vivo data, the recent development of transfer learning techniques may provide a solution for adapting AIDTox trained with in vitro data for in vivo risk prediction.

AUTHOR CONTRIBUTIONS

J.H.M., Y.H., and J.D.R. wrote the manuscript. J.H.M. and Y.H. designed the research. Y.H. and J.D.R. performed the research. J.H.M., Y.H., and J.D.R. analyzed the data.

FUNDING INFORMATION

This work was supported by NIH grants P30 ES013508, R01 LM010098, R01 AG066833, and K99 LM013646.

CONFLICT OF INTEREST STATEMENT

The authors declared no competing interests for this work.

REFERENCES

- Hemmerich J, Ecker GF. In silico toxicology: from structure-activity relationships towards deep learning and adverse outcome pathways. *Wiley Interdisc Rev Comput Mol Sci*. 2020;10:e1475.
- Cherkasov A, Muratov EN, Fourches D, et al. QSAR modeling: where have you been? Where are you going to? *J Med Chem*. 2014;57:4977-5010.
- Richard AM, Huang R, Waidyanatha S, et al. The Tox21 10K compound library: collaborative chemistry advancing toxicology. *Chem Res Toxicol*. 2021;34:189-216.
- Kleinstreuer NC, Yang J, Berg EL, et al. Phenotypic screening of the ToxCast chemical library to classify toxic and therapeutic mechanisms. *Nat Biotechnol*. 2014;32:583-591.
- Sedykh A, Zhu H, Tang H, et al. Use of in vitro HTS-derived concentration-response data as biological descriptors improves the accuracy of QSAR models of in vivo toxicity. *Environ Health Perspect*. 2011;119:364-370.
- Liu J, Mansouri K, Judson RS, et al. Predicting hepatotoxicity using ToxCast in vitro bioactivity and chemical structure. *Chem Res Toxicol*. 2015;28:738-751.
- Ammad-ud-din M, Georgii E, Gonen M, et al. Integrative and personalized QSAR analysis in cancer by kernelized Bayesian matrix factorization. *J Chem Inf Model*. 2014;54:2347-2359.
- Yamane J, Aburatani S, Imanishi S, et al. Prediction of developmental chemical toxicity based on gene networks of human embryonic stem cells. *Nucleic Acids Res*. 2016;44:5515-5528.
- Capuzzi SJ, Politi R, Isayev O, Farag S, Tropsha A. QSAR modeling of Tox21 challenge stress response and nuclear receptor signaling toxicity assays. *Front Environ Sci*. 2016;4:3.
- Zhang J, Mucs D, Norinder U, Svensson F. LightGBM: an effective and scalable algorithm for prediction of chemical toxicity-application to the Tox21 and mutagenicity data sets. *J Chem Inf Model*. 2019;59:4150-4158.
- Mayr A, Klambauer G, Unterthiner T, Hochreiter S. DeepTox: toxicity prediction using deep learning. *Frontiers in environmental. Science*. 2016;3:80.
- Idakwo G, Thangapandian S, Luttrell J, Zhou Z, Zhang C, Gong P. Deep learning-based structure-activity relationship modeling for multi-category toxicity classification: a case study of 10K Tox21 chemicals with high-throughput cell-based androgen receptor bioassay data. *Front Physiol*. 2019;10:1044.
- Matsuzaka Y, Uesawa Y. Molecular image-based prediction models of nuclear receptor agonists and antagonists using the DeepSnap-deep learning approach with the Tox21 10K library. *Molecules*. 2020;25:2764.
- Polishchuk P. Interpretation of quantitative structure-activity relationship models: past, present, and future. *J Chem Inf Model*. 2017;57:2618-2639.
- Ma J, Yu MK, Fong S, et al. Using deep learning to model the hierarchical structure and function of a cell. *Nat Methods*. 2018;15:290-298.
- Elmarakeby HA, Hwang J, Arafeh R, et al. Biologically informed deep neural network for prostate cancer discovery. *Nature*. 2021;598:348-352.
- Lin CH, Lichtarge O. Using interpretable deep learning to model cancer dependencies. *Bioinformatics*. 2021;37:2675-2681.
- Hao Y, Romano JD, Moore JH. Knowledge-guided deep learning models of drug toxicity improve interpretation. *Patterns*. 2022;3:100565.
- Romano JD, Hao Y, Moore JH, Penning TM. Automating predictive toxicology using ComptoxAI. *Chem Res Toxicol*. 2022;35:1370-1382.
- Romano JD, Hao Y, Moore JH. Improving QSAR modeling for predictive toxicology using publicly aggregated semantic graph data and graph neural networks. *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2022*. World Scientific; 2021.
- Urbanowicz RJ, Olson RS, Schmitt P, Meeker M, Moore JH. Benchmarking relief-based feature selection methods for bioinformatics data mining. *J Biomed Inform*. 2018;85:168-188.
- Jassal B, Matthews L, Viteri G, et al. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2020;48:D498-D503.
- Kingma DP, Ba J. Adam: A method for stochastic optimization. 2014. arXiv preprint arXiv:1412.6980.
- Cereto-Massague A, Ojeda MJ, Valls C, Mulero M, Garcia-Vallve S, Pujadas G. Molecular fingerprint similarity search in virtual screening. *Methods*. 2015;71:58-63.
- Quinn NP, García-Gutiérrez L, Doherty C, et al. IQGAP1 is a scaffold of the core proteins of the hippo pathway and negatively regulates the pro-apoptotic signal mediated by this pathway. *Cell*. 2021;10:478.
- Eadie L, Hughes T, White D. Interaction of the efflux transporters ABCB1 and ABCG2 with imatinib, nilotinib, and dasatinib. *Clin Pharmacol Therapeut*. 2014;95:294-306.
- Veith A, Moorthy B. Role of cytochrome P450s in the generation and metabolism of reactive oxygen species. *Curr Opin Toxicol*. 2018;7:44-51.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Hao Y, Romano JD, Moore JH. Knowledge graph aids comprehensive explanation of drug and chemical toxicity. *CPT Pharmacometrics Syst Pharmacol*. 2023;12:1072-1079. doi:[10.1002/psp4.12975](https://doi.org/10.1002/psp4.12975)